Economics Letters 238 (2024) 111676

Contents lists available at ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

On the anatomy of cyberattacks☆

Jin-Wook Chang, Kartik Jayachandran, Carlos A. Ramírez *, Ali Tintera

Federal Reserve Board, United States

ARTICLE INFO

JEL classification:
D39
D22
M15

Keywords:
Cyberattacks
Cybersecurity
Hacks
Cyber risk

ABSTRACT

Using detailed information on cyberattacks and establishments in the United States, we study whether and how an establishment's characteristics can alter the likelihood of cyberattacks. We find that larger establishments and establishments of publicly traded companies are more likely targets.

1. Introduction

The increased frequency and severity of cyberattacks has recently attracted considerable attention. Despite that cybersecurity threats consistently ranked among the top 10 concerns of business, government, and academic leaders, the distribution of cyberattacks across institutions is, at best, imperfectly understood, as public data are scant and mostly anecdotal.[1] This paper partially fills this gap by using detailed information on cyberattacks and establishments in the U.S. and studying which types of establishments are more likely to be targeted.

Ex-ante, it is not obvious which institutions are more likely to become victims of cyberattacks. This is because, from a theoretical perspective, the equilibrium distribution of cyberattacks depends on hackers' motivation and institutions' response, both of which are difficult to observe.[2] To understand this observation more clearly, consider a simple economy wherein hackers are purely financially motivated. And assume hackers target larger institutions with the expectation of obtaining higher ransoms. Considering this strategy, larger institutions would increase their investments in cybersecurity, making it more difficult to implement successful attacks. Thus, on expectation, targeting larger institutions may become less profitable. A similar idea applies to smaller institutions. Here, however, expected profits from successful attacks might be smaller, as smaller institutions might be unable to pay high ransoms. Consequently, hackers might have less incentives to target such institutions to begin with. Due to these forces, hackers might target institutions at random. As a result, the equilibrium distribution of cyberattacks might closely resemble the size distribution of institutions within the economy.[3]

From an empirical perspective, scant data on both cyberattacks and private institutions poses a significant challenge. Besides the lack of public data on cyberattacks, it is difficult to find detailed information on private companies, many of which are themselves victims of cyberattacks. To tackle this challenge, we combine a comprehensive data set on cyberattacks with the National Establishment Time Series (NETS) data set (Walls & Associates, 2020) to provide a representative description of the anatomy of cyberattacks across U.S. institutions.

With these data in hand—which account for about 2.5 million observations at the establishment–year level—we show that establishments of publicly traded companies are 2.68 times more likely to be targeted than establishments of nonpublic institutions—which, within our sample, cover non-publicly traded companies, non-profits, and government

☆ The views expressed herein are ours and might not necessarily reflect those from Federal Reserve Board or other members of its staff.

* Corresponding author at: Federal Reserve Board, 1801 K Street NW, Washington DC, 20006, United States.
E-mail addresses: jin-wook.chang@frb.gov (J.-W. Chang), kartik.jayachandran@frb.gov (K. Jayachandran), carlos.ramirez@frb.gov (C.A. Ramírez), ali.tintera@frb.gov (A. Tintera).

[1] The Global Risk Report of the World Economic Forum consistently reports cybersecurity threats among the top 10 concerns among world economic leaders. For more details, see https://www.weforum.org/publications/series/global-risks-report/. The Cybersecurity and Financial System Resilience Report of the Federal Reserve Board provides a descriptive account of policymakers' concerns about the potential system-wide repercussions of cyberattacks and measures taken to strengthen cybersecurity within the financial sector. For more details, see https://www.federalreserve.gov/publications/cybersecurity-and-financial-system-resilience-report.htm. Kashyap and Wetherilt (2019) emphasize the micro and macroprudential challenges posed by cyberattacks in modern economies.

[2] See Ablon (2018) for a description of hackers' motivations and their different types.

[3] See Dziubinski and Goyal (2013, 2017), Block et al. (2020) and Block et al. (2022) for equilibrium models of attack and defense within a network framework.

https://doi.org/10.1016/j.econlet.2024.111676
Received 20 January 2024; Received in revised form 27 February 2024; Accepted 24 March 2024
Available online 9 April 2024
0165-1765/Published by Elsevier B.V.

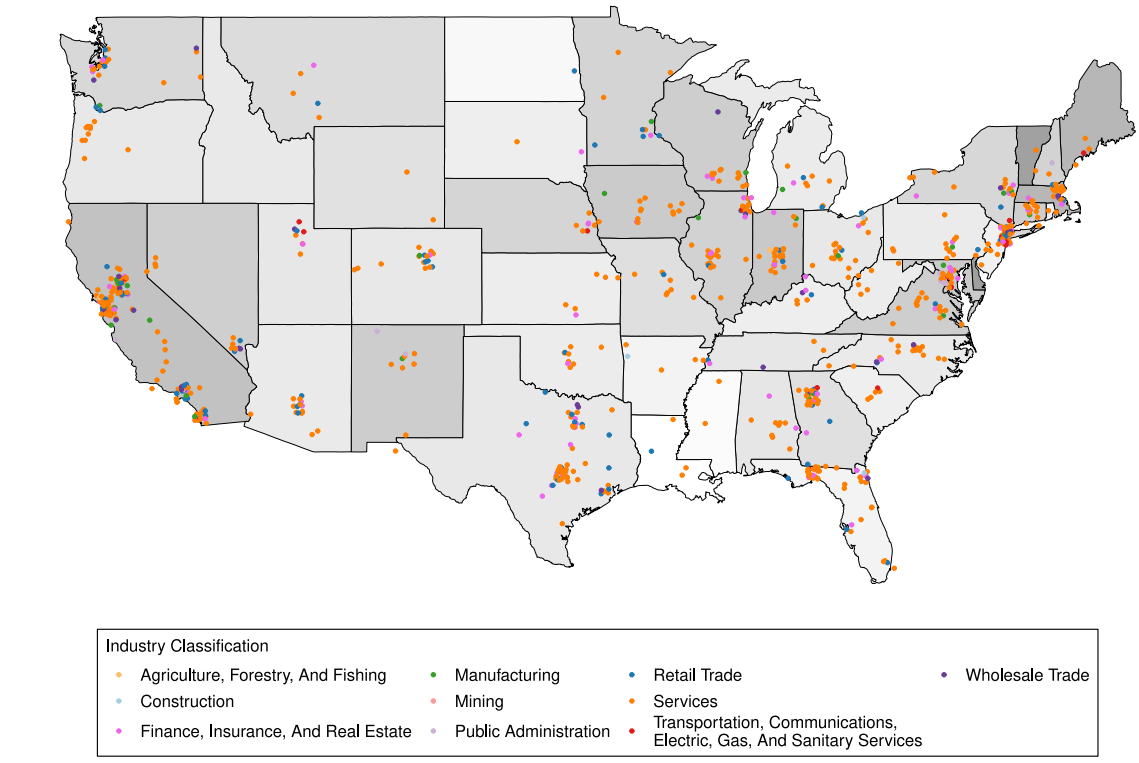

Fig. 1. Distribution of hacks across the U.S. within our sample.

institutions. When compared with the average establishment in our sample, establishments generating 100 million dollars more in annual sales are 9.52% more likely to be targeted. And establishments with 100 more employees are 0.90% more likely to be victims of cyberattacks. Results at the institution level are even more striking. Publicly traded companies are 9.32 times more likely to be targeted than nonpublic institutions. And when compared with the average institution, institutions with 100 more employees are 2.12% more likely to be targeted. Our results control for various fixed-effects and are robust to variation in regression specifications and merging methodologies.

Our findings are consistent with the idea that publicly traded companies and larger institutions are more likely targets of cyberattacks. Our analysis complements previous work examining cyber risk, including Jamilov et al. (2021), Kamiya et al. (2021), and Florackis et al. (2023). Although our paper and this literature share an emphasis on the distribution of cyberattacks, we provide a more granular picture of the anatomy of hacks across the whole distribution of U.S. establishments and not just publicly traded corporations. Our analysis also complements a literature examining the determinants of cyber risk—see, (Aldasoro et al., 2020). Here, we provide a more detailed account of hacks across both public and private U.S. companies. Our results also complement a literature that emphasizes the potential system-wide implications of cyberattacks, including Duffie and Younger (2019), Kashyap and Wetherilt (2019), Kotidis and Schreft (2022), and Eisenbach et al. (2022).

2. Data and summary statistics

To identify cyberattacks we use the Privacy Rights Clearinghouse (PRC) Data Breaches database—a collection of privacy breaches as reported by state Attorney Generals and the U.S. Department of Health and Human Services. Although these data contain over 9000 observations spanning from 2005 through 2019, only a subsample of them affects U.S. institutions.[4] Because we are primarily interested in hacks—defined as breaches caused by an outside party or malware—affecting U.S. institutions, our initial sample contains 2508 observations from 2005 to 2019. Besides institutions' names, observations in PRC provide the geographical location (city and state) of hacked institutions as well as the year of the hack.

Because many observations in PRC refer to non-publicly traded companies and government institutions, we resort to NETS—a representative inventory of U.S. businesses with granular information for almost 80 million (private and public) establishments—to obtain characteristics of hacked institutions within our initial sample.[5] Our merging methodology matches observations in our initial sample with NETS establishments by name and location. We purposely generate our match at the establishment–year level to exploit potential variation across establishments within institutions. This mapping also helps us tackle concerns regarding over-representation of large publicly traded companies accounting for a multitude of establishments across the U.S.

Out of the 2508 initial observations, our matching process generates an intermediate sample of 1220 establishment–year observations for which we have detailed business information from 2005 to 2019.[6] Fig. 1 depicts the geographical distribution of cyberattacks within our

[4] Because public information about cyber incidents is scant, the PRC data have been frequently used as a good approximation of cyber incidents in the U.S.; see Jamilov et al. (2021), Kamiya et al. (2021), and Florackis et al. (2023), among many others.

[5] Barnatchez et al. (2017) find that NETS can be a useful private-sector source of business microdata—relative to official U.S. business universe data sources—for studying business activity in granularity. Importantly, NETS can be accessed without extensive proposal, security clearance processes, and the need to be accessed inside of secure government facilities, potentially providing an efficient way to conduct research when business-level microdata is needed.

[6] Out of the 2508 observations in PRC, 19 observations lack any type of location data, 24 observations do not have state information, and 686

Table 1
Summary statistics.

Panel A: Establishment level							
	Mean	S.D.	10th	25th	50th	75th	90th
# of employees	6.89	13.92	1	2	2	5	15
Annual sales (in millions)	0.69	1.69	0.05	0.08	0.15	0.73	1.4
Ratio of public companies (in %)	2.4	0.31	2.1	2.2	2.3	2.5	2.8
Panel B: Institution level							
	Mean	S.D.	10th	25th	50th	75th	90th
# of employees	7.21	16.57	1	2	2	4	13
Annual sales (in millions)	0.75	2.14	0.05	0.08	0.15	0.33	1.13
Ratio of public companies (in %)	1	0.13	0.8	0.92	0.99	1	1.2

This table reports statistics of establishments and institutions in our baseline sample at the annual frequency. Our sample contains 2,499,369 observations at the establishment–year level from 2005 to 2019. Panel A reports statistics at the establishment level while Panel B reports statistics at the institution level.

intermediate sample. Dots represent the location of hacked establishments. And colors are assigned according to their SIC division. States are colored according to the fraction of their establishments affected by cyberattacks in our sample—the lighter the color, the lower the fraction. Although many hacks in our sample affect establishments in California, Texas, New York, and Florida, Fig. 1 shows that cyberattacks are somewhat equally spread across states. Fig. 1 also shows that hacks affect establishments across a variety of different economic sectors.

Because observing more hacks affecting establishments with a specific characteristic might be just a reflection of the fact that there are more establishments with such characteristic, we combine the above data with a large random sample of NETS establishments. Our idea is to add controls and improve the representativeness of our data, obtaining a better picture of the anatomy of cyberattacks across U.S. establishments. In particular, we add a random sample of about 415,000 NETS establishments to our intermediate data. As a result, we obtain a sample with about 2.5 million establishment–year observations. For each establishment, we retrieve detailed information, including business location, headquarters, employment, sales, and other establishment–level data at the annual frequency, from 2005 to 2019.

Table 1 reports summary statistics of our baseline sample. Panel A reports statistics at the establishment–year level. The average establishment employs a bit less than seven employees per year and generates sales for about 690,000 dollars. On an average year, 2% of establishments belong to a publicly traded company. Panel B reports statistics at the institution–year level. As Panel B shows, the average institution in our sample employs a bit more than seven employees per year and generates sales of around 750,000 dollars. On an average year, 1% of institutions are publicly traded companies. The juxtaposition of Panels A and B shows that most institutions in our sample are small, non-publicly traded, and composed of, at most, one establishment.

3. Empirical approach and results

With our baseline sample in hand, we use the following logistic regression to explore whether an establishment's characteristics can alter the likelihood of being the target of a cyberattack:

$$\log\left(\frac{p_{it}}{1-p_{it}}\right) = \beta' X_{it} + \epsilon_{it}, \tag{1}$$

where there are observations on establishments (i) across years (t). The above equation uncovers a relationship between the logarithmic odds ratio, $\log\left(\frac{p_{it}}{1-p_{it}}\right)$, and establishment i's characteristics—wherein p_{it} captures the likelihood that establishment i is hacked at year t. Here, X_{it} is a vector of explanatory and control variables, which includes the constant term, and ϵ_{it} represents the error term. Explanatory variables include two measures of size—annual sales and number of employees—and whether an establishment belongs to a publicly traded company. To control for unobserved heterogeneity associated with characteristics at the industry-, state-, and year-levels, we include industry-, state-, and year-fixed effects. We also cluster standard errors at the industry-state level to correct for potential autocorrelation among residuals.

Table 2 reports our central findings. Panel A presents results at the establishment level while Panel B reports results at the institution level. For completeness, the first 6 columns in both Panels report different subsets of our explanatory variables while our most robust specifications are reported in columns 7. As Table 2 shows, our explanatory variables are statistically significant across most specifications. Panel A shows that larger establishments and establishments of publicly traded companies are more likely to be targeted. Panel B shows that this result also holds at the institution level.

As column 7 of Panel A shows, the coefficient associated with the public/private dummy (0.98748) is statistically significant. The same applies to the coefficients associated with sales (0.00091) and the number of employees (0.00009). Within a logistic regression framework, these values mean that establishments of publicly traded companies are 2.68 times more likely to be targeted than establishments of government or non-publicly traded institutions. When compared with the average establishment in our sample, establishments generating 100 million more in annual sales are 9.52% more likely to be targeted. And establishments with 100 more employees are 0.90% more likely to be victims of cyberattacks.

Largely consistent with Panel A's findings, Panel B shows that publicly traded companies and larger institutions are more likely to be targets. After converting coefficients to their exponentiated values for interpretation, Panel B shows that publicly traded companies are 9.32 times more likely to be targeted than nonpublic institutions. And, when compared to the average institution, institutions with 100 more employees are 2.12% more likely to be targeted.[7]

3.1. Discussion and data limitations

Taken together, our findings support the view that larger establishments are more likely to become targets of cyberattacks. The same applies to establishments of publicly traded companies. Although many nonpublic and small institutions are frequent targets, not taking into consideration the overall distribution of establishments within the economy can lead to wrong conclusions.

observations do not have city information. And 28% of observations have at least one piece of location data missing where only 1817 observations have name, city, and state information. In our baseline sample, we match 1396 observations in PRC. Hence, considering that only 1817 observations have name, city, and state information, our approximate matching rate is closer to 77%. A more detailed description of our merging methodology appears in Section A of the Online Appendix.

[7] Section C in the Online Appendix shows that our findings are consistent with results from running probit (instead of logit) regressions and robust to variations in our merging methodology.

Table 2
Central findings.

	Dependent variable: $\log(p_{it}/(1-p_{it}))$						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Panel A: Results at the establishment level						
Public/Private dummy	1.02847***			0.98934***	1.01025***		0.98748***
	(0.31518)			(0.31257)	(0.31546)		(0.16251)
Sales		0.00132***		0.00114***		0.00107***	0.00091***
		(0.00021)		(0.00023)		(0.00021)	(0.00022)
# employees			0.00016***		0.00014***	0.00010***	0.00009***
			(0.00003)		(0.00003)	(0.00004)	(0.00002)
Observations	2,341,562	2,341,562	2,341,562	2,341,562	2,341,562	2,341,562	2,341,562
R-squared	0.08684	0.08479	0.08430	0.08766	0.08730	0.08485	0.08771
	Panel B: Results at the institution level						
Public/Private dummy	2.31050***			2.19597***	2.22263***		2.23217***
	(0.33088)			(0.33184)	(0.32441)		(0.17962)
Sales		0.00083***		0.00066***		0.00016	−0.00008
		(0.00021)		(0.00018)		(0.00024)	(0.00014)
# employees			0.00021***		0.00020***	0.00019***	0.00021***
			(0.00002)		(0.00001)	(0.00002)	(0.00003)
Observations	2,242,172	2,242,172	2,242,172	2,242,172	2,242,172	2,242,172	2,242,172
R-squared	0.11154	0.10034	0.10435	0.11435	0.11903	0.10432	0.11894
Controls:							
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

Although PRC is, to the best of our knowledge, the most comprehensive public data on cyberattacks, we are mindful of its selection biases. Small and private companies could be underrepresented as they might not have the technology to uncover cyberattacks or may not bother to report them to authorities. Large and publicly traded companies might also have incentives to underreport—see, for example, (Kamiya et al., 2021). In addition, larger and publicly traded companies might have the resources to hide these incidents. Because the selection bias can potentially go in either direction we do not take a stance on it and use the PRC data as it is.

4. Conclusion

Using granular data on cyberattacks and establishments in the United States we study whether and how an institution's characteristics can alter the likelihood of being the target of cyberattacks. We find that larger establishments—in terms of sales and number of employees—and establishments of publicly traded companies are more likely targets. A similar result holds at the institution level. Our results are robust to variation in regression specifications and merging methodologies.

Declaration of competing interest

There are no competing interests to declare.

Data availability

The authors do not have permission to share data.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.econlet.2024.111676.

References

Ablon, Lillian, 2018. Data thieves: The motivations of cyber threat actors and their use and monetization of stolen data. Testimony presented before the House Financial Services Committee, Subcommittee on Terrorism, and Illicit Finance, on March 15, 2018.

Aldasoro, Iñaki, Gambacorta, Leonardo, Giudici, Paolo, Leach, Thomas, 2020. The Drivers of Cyber Risk. BIS Working Paper, (865).

Barnatchez, Keith, Crane, Leland D., Decker, Ryan A., 2017. An Assessment of National Establishment Time Series (NETS) Database. Finance and Economics Discussion Series (FEDS), (110).

Block, Francis, Chatterjee, Kalyan, Dutta, Bhaskar, 2022. Attack and interception in networks. Theor. Econ. 18, 1511–1546.

Block, Francis, Dutta, Bhaskar, Dziubinski, Marcin, 2020. A game of hide and seek in networks. J. Econom. Theory 190.

Duffie, Darrell, Younger, Joshua, 2019. Cyber Runs: How a Cyber Attack Could Affect U.S. Financial Institutions. Hutchins Center Working Paper, Vol. 51.

Dziubinski, Marcin, Goyal, Sanjeev, 2013. Network design and defense. Games Econom. Behav. 79, 30–43.

Dziubinski, Marcin, Goyal, Sanjeev, 2017. How to defend a network? Theor. Econ. 12, 331–376.

Eisenbach, Thomas M., Kovner, Anna, Lee, Michael Junho, 2022. Cyber risk and the U.S. financial system: A pre-mortem analysis. J. Financ. Econ. 145, 802–826.

Florackis, Chris, Louca, Christodoulos, Michaely, Roni, Weber, Michael, 2023. Cybersecurity risk. Rev. Financ. Stud. 36, 351–407.

Jamilov, Rustam, Rey, Hélene, Tahoun, Ahmed, 2021. The Anatomy of Cyber Risk. NBER Working Paper Series, (28906).

Kamiya, Shinichi, Kang, Jun-Koo, Kim, Jungmin, Milidonis, Andreas, Stulz, Rene M., 2021. Risk management, firm reputation, and the impact of successful cyberattacks on target firms. J. Financ. Econ. 139, 719–749.

Kashyap, Anil K., Wetherilt, Anne, 2019. Some principles for regulating cyber risk. In: AEA Papers and Proceedings. Vol. 109, pp. 482–487.

Kotidis, Antonis, Schreft, Stacey L., 2022. Cyberattacks and Financial Stability: Evidence from a Natural Experiment. Finance and Economics Discussion Series (FEDS), (25).

Walls & Associates, 2020. National establishment time-series (NETS) database.